Chapter 12

Estimating the Prevalence and Correlates of Serious Mental Illness in Community Epidemiological Surveys

Ronald C. Kessler, Ph.D.
Department of Health Care Policy, Harvard Medical School

Patricia A. Berglund, M.B.A.
Institute for Social Research, University of Michigan

Meyer D. Glantz, Ph.D.
Division of Epidemiology, Services, and Prevention Research, National Institute of Drug Abuse

Doreen S. Koretz, Ph.D.
Division of Mental Disorders, Behavioral Research and AIDS, National Institute of Mental Health

Kathleen R. Merikangas, Ph.D.
Mood and Anxiety Disorders Program, National Institute of Mental Health

Ellen E. Walters, M.S.
Alan M. Zaslavsky, Ph.D.
Department of Health Care Policy, Harvard Medical School

Introduction

Public Law (PL) 102-321, the Alcohol, Drug Abuse, and Mental Health Administration (ADAMHA) Reorganization Act, established a block grant for States to fund community mental health services for adults with serious mental illness (SMI) and children with severe emotional disturbances (SED). The law required States to include SMI incidence and prevalence estimates in their annual applications for block grant funds. The law also required the Substance Abuse and Mental Health Services Administration (SAMHSA) to develop an operational definition of SMI and to create an estimation methodology based on this definition for use by States. The definition of SMI in PL 102-321 requires an individual to have at least one 12-month *Diagnostic and Statistical Manual* (DSM) disorder other than a substance use disorder (SUD), and to have "serious impairment." SAMHSA subsequently decided that "serious impairment" should be defined as a Global Assessment of Functioning (GAF) score below 60 (Endicott, et al., 1976; SAMHSA, 1993).

SAMHSA established a task force to develop a methodology to estimate the incidence and prevalence of SMI. The first step was to reanalyze data from the two recent major psychiatric epidemiological surveys of DSM disorders in the United States, the Epidemiological Catchment Area (ECA) study (Robins and Regier, 1991), and the National Comorbidity Survey (NCS) (Kessler et al., 1994). The goals of the task force were to estimate the prevalence of SMI in the country as a whole and to examine sociodemographic correlates of SMI. This work was carried out with the recognition that neither the ECA nor the NCS was designed with the goal of estimating SMI. Imprecise *post hoc* indicators of impair-

The developmental research for the CIDI-SF scales was supported by National Institute of Mental Health (NIMH) grants MH46376, MH49098, and MH52861, with supplemental support from the W.T. Grant Foundation (grant 90135190). The developmental research for the K10/K6 scales was supported by the John D. and Catherine T. MacArthur Foundation Network on Successful Midlife Development (Gilbert Brim, director). The SMI screening scale project was supported by Contract 283-98-9008 from SAMHSA. The implementation of K10/K6 calibration in the new NCS-R is supported by NIMH Collaborative Agreement U01-MH60220 with supplemental support from the National Institute of Drug Abuse and the Robert Wood Johnson Foundation (grant 044780). In addition to the above grants and contracts, preparation of this report was supported by NIMH grant MH066627.

ment consequently had to be used to approximate the SAMSHA definition. On the basis of this work, the task force estimated that 5.4 percent of the U.S. noninstitutionalized civilian population meet criteria for SMI at some time each year (Kessler et al., 1996). The task force found that SMI is significantly more common among women than men, more common among previously married and never married people than among married people, and inversely related to both income and education (Kessler et al., 1999). SMI was found to be insignificantly related to race/ethnicity, living in an urban setting, and region of the country.

The second step for the task force was to explore whether precise estimates of the prevalence of SMI could be generated for counties and States by applying standard, small-area estimation methods to the ECA and NCS data (Schaible, 1996). This exploration showed clearly that precise estimates of this sort cannot be made currently because of the weak associations between available small-area demographic variables and SMI (Kessler et al., 1999). The task force began its exploration by developing prediction equations for SMI in the NCS data that exclusively used predictors available for each county and State through the Area Resources File (ARF). The ARF is a compendium of data on a wide range of topics (e.g., demography, topography, weather conditions, road and traffic conditions, health care resources, criminal justice) that is assembled from a variety of government sources and continually updated. ARF information was applied to these equations to generate a predicted prevalence of SMI for each county and State in the country. Appropriate standard errors of these estimates were then generated to adjust for the imprecision of the prediction equations. The resulting estimates were shown not to differ significantly across counties and States. Although this result could have occurred because the true prevalence of SMI is the same in all counties and States in the United States, a more plausible interpretation, and the one adopted by the SMI task force, is that the prediction equations were too weak and the geographic variation in these predictors was too small to detect the true differences in the prevalence of SMI across counties and States.

A Screening Scale for SMI

On the basis of these results, the task force recommended that an SMI screening scale be developed for use in ongoing government surveys, such as the National Health Interview Survey (NHIS) and the SAMHSA National Household Survey on Drug Abuse (NHSDA), and that analysis of these results be used to generate small-area estimates of SMI for counties and States. SAMHSA funded a methodological study to implement this recommendation. This study considered three measures as possible screens for SMI. The first was a truncated version of the World Health Organization (WHO) Composite International Diagnostic Interview Short-Form (CIDI-SF) scales (Kessler, et al., 1998). These disorder-specific scales are designed to assign predicted probabilities of meeting 12-month criteria for a number of DSM-IV anxiety and mood disorders on the basis of a short series of questions for each disorder.

The second measure was a modified version of the K10/K6 scales of nonspecific psychological distress (Kessler et al., in press). These short (6 and 10 questions) scales were developed for use in the core of the redesigned U.S. NHIS to measure the frequency of commonly occurring symptoms of psychological distress (e.g., worry, restlessness, sadness) over a 30-day recall period. The K10/K6 scales were modified for use in the SAMHSA methodological study to ask about symptoms in the month in the past year when the respondent's emotional problems were worst.

The third measure was a truncated version of the WHO Disability Assessment Schedule (WHO-DAS) (Rehm et al., 1999). The WHO-DAS was developed to operationalize the core dimensions in the WHO International Classification of Functioning, Disability, and Health (World Health Organization, 2001) by asking about the severity of illness-related impairments in a variety of role domains over a 30-day recall period. The WHO-DAS was modified for use in the SAMHSA methodological study to ask about impairments caused by emotional problems during the month in the past year when the respondent's emotional problems were worst.

All three scales were administered to the second-stage sample of a two-stage general population convenience sample. The first stage of this sample consisted of 1,000 people who were screened by telephone for serious mental health problems. The second stage consisted of 155 respondents selected from the first-stage sample to oversample people with suspected serious mental health problems. Second-stage respondents were interviewed face to face in their homes by trained clinical interviewers. The interviews began with respondents self-administering the CIDI-SF, K10/K6, and WHO-DAS scales

in a way that kept responses hidden from the interviewer. The interviewer then administered the Structured Clinical Interview for DSM-IV (SCID) (First, et al., 1997) and rated respondents on the GAF while remaining blind to the responses on the screening scales. Respondents were classified as having SMI or not on the basis of their SCID and GAF ratings. Logistic regression analysis was then used to estimate associations between the screening scales and SMI. As reported previously (Kessler et al., 2003), the K6 was found to be the most powerful predictor of SMI, with an area under the Receiver Operator Characteristic Curve of .86. None of the other screening scales significantly predicted SMI after scores on the K6 were controlled. The optimal cut-point on the K6 in terms of equalizing false positives and false negatives in the weighted second-stage sample was 0 to 12 versus 13+, coding item responses 0 to 4 and summing items to yield a scale with a 0 to 24 range. Sensitivity (standard error in parentheses) was .36 (.08), specificity was .96 (.02), and total classification accuracy was .92 (.02) at this cut-point.

Scoring the Screening Scale

On the basis of these results, the K6 was added to the NHSDA in 2001. Results are not yet available. However, as noted, the K6 was originally developed for use in the NHIS, and it has been included in the NHIS since 1997. With a random adult sample of more than 40,000 respondents each year, K6 reports have now been obtained on approximately a quarter of a million people in the NHIS. Given the strength of the association between the K6 and SMI in the SAMHSA methodological study, this large number of nationally representative cases should be enough to generate fairly precise estimates of the prevalence of SMI in most States. However, doing so requires calibration rules to be available. The SAMHSA methodological study was too small and too unsystematic to generate such rules. This statement might be confusing to readers in light of the fact, noted above, that the optimal cut-point in terms of balancing false positives with false negatives in the methodological study was between the range 0 to 12 (predicted noncases) and the range 13 to 24 (predicted cases) on the 0 to 24 range of the K6. A superficial way to code the K6 in the NHIS and NHSDA surveys would be to use this rule to define likely cases. However, it is important to recognize that this cut-point is only an approximation and is only "optimal," even in the narrowly defined sense proposed here (i.e., balancing false positives and false negatives), when the probability of SMI at a given score on the K6 (positive predictive value) is constant across samples. Positive predictive value will not be constant, though, even if the separate conditional distributions of K6 scores among people with SMI (sensitivity) and without SMI (specificity) are constant across samples, unless the true prevalence of SMI remains constant. The reason is that any deviation in the proportion of people with actual SMI in the new samples will lead to changes in positive predictive value at a given level of the K6 (Goldberg, Oldehinkel, and Ormel, 1998). As a result, specifying a single threshold for SMI on the K6 that is applied to all samples is not appropriate.

This problem can be solved, but doing so requires access to a much larger calibration sample than the 155 cases included in the SAMHSA methodological study. When such a sample is available, scoring rules can be developed that allow predicted probabilities of SMI to be estimated for respondents in other samples, such as the NHIS or NHSDA, on the basis of the assumption of consistent sensitivity and specificity of the K6 across samples. This assumption is much more plausible than the assumption of consistent positive and negative predictive values. The assumption of consistent sensitivity and specificity can be made either for an entire sample or for important subsamples (e.g., gender, age, educational attainment). In the ideal case, subsample differences in the calibration sample should be analyzed to evaluate the plausibility of assuming that sensitivity and specificity are constant across subsamples and, when this assumption is rejected, to estimate separate sensitivity and specificity values in informative subsamples (Furukawa, et al., 2003).

The assumption of sensitivity and specificity being constant across samples is equivalent to the assumption that stratum-specific likelihood ratios (SSLRs) are constant, but the parameterization of the assumption in terms of SSLRs has computational advantages over the parameterization in terms of sensitivities and specificities. An SSLR is an odds ratio (OR) that compares respondents who have a specific score on a screening scale (in this case, the K6) with those who have all other scores on the scale in terms of their odds of having a dichotomous outcome (in this case, SMI) (Guyatt and Rennie, 2001). Once the SSLRs are calculated, Bayes' theorem can be used to show that

$$\text{POO} \times \text{SSLR} = \text{ROO}, \qquad (1)$$

where POO is the population odds of the dichotomous outcome (which can be calculated from the population prevalence) and ROO is the individual respondent's odds of the outcome. The individual's probability of the dichotomous outcome, p, can easily be derived from ROO by using the transformation

$$\text{ROO} = p / (1 - p). \quad (2)$$

The results in equations (1) and (2) can be used to assign individual-level predicted probabilities of SMI based on individual K6 scores, but only after knowing the prevalence of SMI in the population. In other words, equations (1) and (2) make it clear that the probability of SMI for any one individual with a given score on the K6 scale varies with the prevalence of SMI in the population from which that individual was drawn. The real problem, then, is to estimate the aggregate prevalence. Fortunately, this can be done using a method that avoids the problem of having to assume constant positive predictive value. This method uses maximum likelihood to compare the empirical K6 distribution in the sample under consideration with the theoretical distributions generated by the sensitivities and specificities in some calibration sample applied to all possible hypothetical prevalences of SMI. The maximum-likelihood estimate of SMI is the one associated with the theoretical distribution of K6 scores most similar to the empirical distribution in the sample. Once this prevalence estimate is obtained, it can be used in equations (1) and (2) to generate individual-level probabilities from individual-level K6 scores for purposes of more detailed analyses of the correlates of SMI. The latter can include the estimation of prevalences in counties and States.

We are unable to implement this optimal scoring approach because the SAMHSA methodological study sample is too small to allow the K6 prevalence to be estimated accurately in the NHIS. However, this problem is being solved by including the K6 in the National Comorbidity Survey Replication (NCS-R) (Kessler and Walters, 2003), a nationally representative face-to-face general population survey of 10,000 respondents that was carried out in 2001–2002. The full NCS-R sample will be used as a calibration sample to generate accurate estimates of the prevalence of SMI from K6 distributions in the NHIS and NHSDA. Scoring rules, as soon as they are available, will be posted on the NCS Web site (http://www.hcp.med.harvard.edu/ncs/).

The diagnostic interview in the NCS-R is a modification of the WHO CIDI (Robins et al., 1988), a fully structured diagnostic interview that assesses a number of commonly occurring mental disorders according to the definitions and criteria of DSM-IV. Respondents are classified as having SMI if they met criteria for any qualifying DSM-IV mental disorder during the year before the interview and if they either scored less than 60 on a structured version of the GAF scale or some equivalent indicator of serious impairment associated with their mental illness during the past year. Because the NCS-R sample is fairly large, it will also be used to explore whether the SSLRs for the relationship between the K6 and SMI are stable across important sociodemographic subsamples and, if not, to generate separate sets of SSLRs in informative subsamples, as well as to collapse K6 scores in ways that optimize the stability of imputations (Peirce and Cornell, 1993). A computer program that generates both an aggregate estimate of the prevalence of SMI using maximum likelihood and individual estimates of the probability of having SMI on the basis of equations (1) and (2) will be posted on the NCS Web site as soon as the NCS-R data are ready for analysis and the K6 calibration has been completed.

Preliminary SMI Prevalence Estimates

Even before the optimal scoring approach is worked out from the NCS-R, it is possible to present very preliminary aggregate estimates of the prevalence of SMI from the NHIS using the simple classification rule that K6 scores in the range 13 to 24 represent likely SMI. On the basis of this rule, the estimated 30-day prevalence of likely SMI in the NHIS is 3.3 percent in 1997, 3.0 percent in 1998, 2.4 percent in 1999, and 2.7 percent in 2000. With a sample of more than 40,000 respondents each year, the standard errors of these annual estimates are less than one-tenth of one percent. This means that the 0.9 percent range of year-to-year variation in the prevalence of SMI over these years is reliable. On a population base of approximately 209 million adults ages 18 and older in the United States, these prevalence estimates are equivalent to an average of between 5.0 million and 6.9 million Americans who met criteria for SMI in any given month of these years.

It would be possible to generate more finely disaggregated NHIS time-trend prevalence estimates to see if there is seasonal variation in SMI, because the NHIS is made up of separate nationally representative samples of about 800 interviews with

adult respondents each week of the year. It would be useful to link these fine-grained time trends to information about current events (e.g., stock market trends, extreme weather patterns, widely publicized world events that might lead to anxiety or depression) in an effort to get insight into reasons for the statistically significant changes in the prevalence of likely SMI over the years 1997 to 2000. Given that the NCS-R calibration rules for transforming K6 scores into estimated probabilities of SMI will soon be available, this analysis of detailed time trends should probably be based on these calibrations rather than on the 0 to 12 versus 13 to 24 dichotomous scoring of the K6.

As noted above, the K6 was modified in the SAMHSA methodological study to ask about symptoms in the month in the past year when the respondent's emotions were worst rather than in the past 30 days. This was done because PL 1021321 defines SMI as a 12-month disorder, and the goal of the methodological study was to screen for the presence of SMI at any time in the year before the interview. The NHSDA adopted this K6 modification. We should consequently expect that the prevalence estimates of SMI in the NHSDA, when they become available, will be higher than in the NHIS because the NHSDA will be estimating 12-month prevalence, while the NHIS estimates 30-day prevalence. A glimpse into this prediction is provided by preliminary data from the first half sample of 5,000 respondents in the NCS-R, 3,015 of whom were administered the K6 questions in exactly the same way as in the NHSDA. The estimated 12-month prevalence of likely SMI in this preliminary sample is 7.2 percent, with a standard error of 0.5 percent. As expected, this is considerably higher than the 2.4 to 3.3 percent 30-day prevalence estimates in the NHIS. On a base of 209 million Americans ages 18 and older, the 12-month SMI prevalence estimate in the preliminary NCS-R data is equivalent to approximately 15 million Americans who met criteria for SMI at some time in the year prior to their participation in the survey.

A decomposition of the preliminary NCS-R SMI data by age and sex is shown in the first column of table 1. Twelve-month SMI is estimated to be more prevalent among women than men (8.6 percent vs. 5.6 percent; $z = 3.25$, $p = 0.001$). Among men, 12-month SMI is estimated to be most prevalent in the age range of 30 to 44 (7.2 percent) and least prevalent in the age range of 60 and older (3.0 percent). Among women, 12-month SMI is estimated to be more prevalent in the age range of 18 to 44 (10.2 to 10.6 percent) and, like men, least prevalent in the age range of 60 and older (4.8 percent). Because of the great deal of interest in the comorbidity between SMI and SUDs, table 1 also shows preliminary estimates for the 12-month prevalence of DSM-IV alcohol or drug abuse or dependence. These SUDs have an estimated 12-month prevalence of 5.0 percent in the total sample, with a standard error of 0.4 percent. It is noteworthy that the DSM-IV criteria for SUD are more restrictive than the criteria in earlier versions of the DSM. This fact partly explains why the estimated prevalence of SUD is lower in the NCS-R than in either the ECA (Robins and Regier, 1991), which was based on DSM-III criteria, or the NCS (Kessler et al., 1994), which was based on DSM-III-R criteria. As in these earlier surveys, though, the 12-month prevalence of SUD is estimated to be much higher among men than women (7.7 percent vs. 2.5 percent; $z = 6.4$, $p < 0.001$). Also consistent with earlier surveys is the strong inverse relationship between age and SUD in the NCS-R. The prevalence range of SUD among NCS-R men is from 16.2 percent in the youngest age group (ages 18 to 29) to 0.2 percent in the oldest age group (60 and older). The comparable range among women is from 6.3 percent in the youngest age group to 0.2 percent in the oldest age group.

If SMI and SUDs were totally unrelated to each other, the prevalence of comorbid SMI and SUD would be the product of the prevalences of the two individual disorders. For example, with total-sample prevalences of 7.2 and 5.0 percent for SMI and SUD, respectively, we would expect that 0.4 percent (.072 x .05) of the population would meet criteria for both SMI and SUD. The actual estimated prevalence of comorbid SMI-SUD is 0.8 percent—twice as high as the prediction based on independence. This lack of independence is consistent with the findings of previous research (Kessler et al., 1996). The last column of table 1 shows ORs between 12-month SMI and 12-month SUD. All the ORs are greater than 1.0, indicating that the prevalence of comorbid SMI-SUD is consistently larger than the product of the component prevalences. Overall, 10.6 percent of respondents with likely SMI also have SUD, compared with 4.6 percent of respondents without SMI. Among respondents with SUD, 18.1 percent of those with substance dependence also have SMI, compared with 12.4 percent of those with substance abuse. Among those without any 12-month SUD, 6.8 percent have SMI.

Table 1. The 12-month prevalences of serious mental illness (SMI), substance use disorder (SUD), and comorbid SMI-SUD in the preliminary national comorbidity survey replication (n = 3,015)

		SMI		SUD		Comorbid SMI-SUD				
Gender	**Age**	**%**	**(se)**	**%**	**(se)**	**%**	**(se)**	**OR**		**(95% CI)**
Male	18–29	5.6	(1.3)	16.2	(2.1)	2.4	(0.9)	4.6	*	(1.8 – 11.6)
	30–44	7.2	(1.3)	8.5	(1.3)	0.7	(0.4)	1.1		(0.3 – 3.9)
	45–59	5.7	(1.2)	3.6	(1.0)	3.9	(0.5)	5.5	*	(1.5 – 20.7)
	60+	3.0	(1.3)	0.2	(0.3)	0.2	(0.3)			—
	Total	5.6	(0.6)	7.7	(0.7)	1.1	(0.3)	3.1	*	(1.7 – 5.7)
Female	18–29	10.6	(1.6)	6.3	(1.2)	1.0	(0.5)	1.8		(0.6 – 5.6)
	30–44	10.2	(1.3)	2.3	(0.6)	0.7	(0.3)	3.8		(0.9 – 15.2)
	45–59	8.5	(1.3)	1.2	(0.5)	0.2	(0.2)	1.7		(0.1 – 22.9)
	60+	4.8	(1.2)	0.2	(0.2)	0.0	0.0			—
	Total	8.6	(0.7)	2.5	(0.4)	0.5	(0.2)	2.6	*	(1.1 – 5.9)
All	18–29	8.1	(1.0)	11.3	(1.2)	1.8	(0.5)	2.4	*	(1.2 – 4.7)
	30–44	8.7	(0.9)	5.4	(0.7)	0.7	(0.3)	1.5		(0.6 – 3.7)
	45–59	7.1	(0.9)	2.4	(0.5)	0.5	(0.2)	3.5	*	(1.1 – 11.0)
	60+	4.1	(0.9)	0.2	(0.2)	0.1	(0.1)	17.2		(0.5 – 637.2)
	Total	7.2	(0.5)	5.0	(0.4)	0.8	(0.2)	2.5	*	(1.6 – 4.0)

Sociodemographic Correlates of SMI, Substance Disorders, and Their Comorbidity

Table 2 shows the results of logistic regression analysis, in which we examined basic sociodemographic predictors of the outcomes in table 1. The first four rows show the age associations in a somewhat different way than in table 1. Significant relationships with age are found both for SMI and SUD. In the case of SMI, the highest prevalence is in middle age, whereas for SUD, the highest prevalence is in young age. The age gradient is also much stronger for SUD than for SMI, leading to an inverse relationship between age and comorbid SMI-SUD. The next entries in the table are for sex. Women are significantly more likely than men to meet criteria for likely SMI, but significantly less likely than men to meet criteria for SUD. The negative association with SUD is stronger than the positive association with SMI, leading to a negative, although nonsignificant, association with comorbid SMI-SUD.

The next entries in table 2 are for education, which is inversely related to SMI, to SUD, and to comorbid SMI-SUD. The associations are essentially multiplicative in predicting comorbid SMI-SUD. That is, the product of the ORs in predicting SMI and SUD is roughly equivalent to the ORs in predicting comorbid SMI-SUD. A rather different pattern is found in the next entries in the table, which deal with marital status. Married people consistently have the lowest rates of disorder, with separated-divorced and never married people having elevated rates of SMI and these same two groups plus cohabiters having elevated rates of SUD. The ORs in predicting comorbid SMI-SUD are roughly multiplicative for the never married, but for cohabiters and the separated-divorced they deviate substantially from this pattern. In the case of cohabiters, there is no elevated risk of comorbid SMI-SUD (OR = 1.3; 95 percent confidence interval = 0.1–17.5), despite the fact that cohabiters have significantly elevated rates of both SMI and SUD. In the case of the separated-divorced, the risk of comorbid SMI-SUD (OR = 17.9; 95 percent confidence interval = 4.0–80.0) is

Table 2. Multiple logistic regressions of 12-month SMI, SUD, and comorbid SMI-SUD on sociodemographic predictors in the preliminary national comorbidity survey replication (n = 3,015)

		SMI		SUD		Comorbid SMI-SUD	
Predictor	**Values**	**OR**	**(95% CI)**	**OR**	**(95% CI)**	**OR**	**(95% CI)**
Age	**18–29**	1.8 *	(1.0 – 3.4)	34.1	(5.3 – 217.7)	15.5	(0.8 – 296.6)
	30–44	2.7 *	(1.6 – 4.6)	23.8	(3.8 – 149.4)	8.7	(0.5 – 158.9)
	45–59	2.3 *	(1.3 – 3.9)	11.5	(1.8 – 74.0)	6.8	(0.4 – 129.3)
	60+	1.0	—	1.0	—	1.0	—
	χ^2_3		15.4*		23.3*		3.9
Gender	**Female**	1.6 *	(1.2 – 2.2)	0.3 *	(0.2 – 0.5)	0.5	(0.2 – 1.3)
	Male	1.0	—	1.0	—	1.0	—
	χ^2_1		11.2*		30.7		2.1
Education	**0–11**	3.4 *	(2.0 – 5.7)	2.7 *	(1.5 – 4.9)	12.8 *	(1.8 – 90.7)
	12	2.7 *	(1.7 – 4.3)	1.7 *	(1.0 – 2.9)	4.2	(0.6 – 29.3)
	13–15	1.9 *	(1.2 – 3.1)	1.5	(0.8 – 2.5)	2.3	(0.3 – 17.7)
	16+	1.0	—	1.0	—	1.0	—
	χ^2_3		25.1		11.3*		12.6*
Marital Status	**Cohabitating**	1.8 *	(1.0 – 3.1)	3.7 *	(2.1 – 6.7)	1.3	(0.1 – 17.5)
	Separated/Divorced	3.2 *	(2.2 – 4.7)	2.8 *	(1.5 – 5.2)	17.9 *	(4.0 – 80.0)
	Widowed	1.7	(0.9 – 3.5)	1.4	(0.2 – 9.8)	—	—
	Never Married	2.3 *	(1.5 – 3.5)	3.2 *	(2.0 – 5.3)	7.2 *	(1.5 – 35.5)
	Married	1.0	—	1.0	—	1.0	—
	χ^2_4		38.0*		28.8*		16.9*
Race/ Ethnicity	**Hispanic**	0.7	(0.5 – 1.2)	0.8	(0.4 – 1.3)	0.8	(0.2 – 3.5)
	Non-Hispanic Black	1.0	(0.6 – 1.5)	0.4 *	(0.2 – 0.8)	0.4	(0.1 – 2.1)
	Non- Hispanic White	1.0	—	1.0	—	1.0	—
	Other	0.9	(0.4 – 1.7)	0.9	(0.4 – 2.0)	1.2	(0.2 – 5.8)
	χ^2_3		1.8		7.3		1.3
Region	**Northeast**	0.7	(0.5 – 1.1)	0.8	(0.4 – 1.3)	3.2	(0.6 – 17.2)
	Midwest	1.0	(0.6 – 1.5)	1.1	(0.7 – 1.9)	4.1	(0.8 – 19.9)
	South	1.0	(0.7 – 1.5)	0.8	(0.5 – 1.3)	1.8	(0.4 – 9.0)
	West	1.0	—	1.0	—	1.0	—
	χ^2_3		4.0		2.8		3.9
Urbanicity	**Major Metro**	1.4	(0.9 – 2.2)	0.8	(0.5 – 1.3)	1.1	(0.4 – 3.8)
	Other Metro	1.3	(0.7 – 2.5)	0.4	(0.2 – 1.0)	0.7	(0.1 – 6.4)
	Rural	1.0	—	1.0	—	1.0	—
	χ^2_2		2.1		3.9		0.3
	$\chi^2_{18/19}$		94.5*		140.7*		42.5*

substantially higher than the product of the ORs for SMI and SUD.

The remaining entries in table 2 show that SMI is not significantly related either to race-ethnicity, region, or urbanicity net of the effects of the other predictors. The failure to find an effect of race-ethnicity is striking in light of concerns among health care policy researchers about health disparities. This finding might be taken to suggest that no such disparities exist with regard to serious emotional or substance problems among racial and ethnic minorities after adjusting for the other variables in this model. It is important to note, however, that previous research has sometimes found interactions between socioeconomic status (e.g., education, income) and race-ethnicity in predicting mental health problems (House, 2002). It may be that the simple additive model estimated here, in which the effects of education are assumed to be the same for people in all race-ethnicity subsamples, masks important effects of minority status that are confined to the sector of the population with low education.

A similar caution can be raised about the failure to find an effect of urbanicity in predicting any of the outcomes in table 2. Considerable concern exists about rural mental health (Hartley et al., 2002), and especially about the possibility that unmet need for treatment is higher in rural areas than urban areas because of low access to specialty care (Bull, Krout, Rathbone-McCuan, and Shreffler, 2001). Our failure to find evidence of significantly elevated SMI or SUD in rural areas is based on a simple additive prediction equation. Perhaps the rural poor or rural minorities have elevated rates of these disorders that could be detected in more detailed analyses. Such analyses were not carried out with the NCS-R data because of low statistical power in the preliminary half-sample data. These analyses are needed, however, when the full NCS-R data set is available. In addition, much more precise data for carrying out these subsample analyses will soon be available in the NHIS and NHSDA once K6 calibration rules are developed on the basis of the full NCS-R data set.

Needs Assessment for Counties

K6 calibration rules based on the full NCS-R data set will be available late in 2003. It will become possible at that time to apply these rules to the data from the NHIS and NHSDA, as well as to other surveys in which the assumption of consistency of SSLRs is plausible. Given what we know of the strength of sociodemographic correlations of SMI and inter-county variation in these correlates, the precision of these estimates will almost certainly come largely from the direct aspect of the estimates; that is, from the number of cases with K6 scores that were assessed in the separate counties and States included in the samples. The number of respondents in even the smallest States will be large enough to generate stable State-level estimates of the prevalence of SMI. For example, with a sample of approximately 250,000 respondents (which could be obtained by pooling the NHIS over 6 years, pooling the NHSDA over 4 years, or pooling and blending the samples from these two surveys over 2 to 3 years), a subsample of approximately 500 respondents will be included from each of the three least populated States (Alaska, Vermont, Wyoming). This is a large enough sample to generate a fairly stable estimate of the prevalence of SMI from K6 scores. However, the sample will be too small to estimate the prevalence of SMI in counties. For example, there will be fewer than 50 respondents in the pooled sample for a county of 50,000 inhabitants. Close to one-fifth of all Americans live in counties with fewer than 50,000 inhabitants.

Mental health needs assessment data are needed at the county level to allow States and counties to make rational resource allocation decisions. One way to obtain such data would be for the Behavioral Risk Factor Surveillance Survey (BRFSS), which is carried out with coordination from the Centers for Disease Control (CDC) every year in each State, to include the K6. BRFSS samples typically include at least 1,500 respondents per State per year. With a sample of this size, it would be possible to generate useful SMI prevalence estimates for groups of counties based on pooled BRFSS data collected over several years. However, the BRFSS currently does not include the K6. The BRFSS includes a small number of mental health screening questions. These questions were evaluated in an early phase of the SAMHSA methodological study and found to be of little value. It would be very useful for the K6 to be substituted for these questions in the BRFSS.

Even the inclusion of the K6 in the BRFSS, however, would not give individual counties the detailed data they need to make informed resource allocation decisions. An intriguing possibility for collecting such county-level data on SMI is to carry out a low-tech county-level telephone needs assessment survey based on the design and methods used in volunteer political polls for local and State elections. In the latter, pollsters who specialize in carrying out local and State election surveys throughout the country develop a preset telephone interview sched-

ule that is known to cover the important issues for such elections, with the possible addition of modest modifications for a particular survey. This interview is programmed onto laptop computers that are shipped to local areas. Local volunteers are then trained to carry out the interview on these computers. The interviews are usually conducted at a temporary phone survey facility (e.g., a church basement, a community center) over a period of 1 or 2 weeks. A team of 20 to 40 volunteer telephone interviewers, each working 10 to 20 hours, can usually complete 500 to 800 telephone interviews over this period. The laptop computers, still containing the completed survey data, are then shipped back to the pollster and a report is prepared.

There is no reason why exactly the same approach could not be used to carry out inexpensive county-level mental health needs assessment surveys. These surveys could include the K6 to screen for SMI along with brief screens for other issues of interest to community mental health planners (e.g., substance problems, suicidality, relationship violence) and questions about barriers to care. A local civic or church group could be recruited to carry out the survey as a voluntary activity. The local telephone company could be asked to donate the cost of installing a temporary telephone bank for the time it would take to carry out the survey. The survey could be repeated every 5 years or so to monitor trends in the prevalence of SMI and service needs for people with SMI.

Results of such surveys could be of great value to county mental health service planners. It would not be difficult to create an interview schedule for these surveys, a laptop computer program that could be used to record interview data, or a standard electronic report format to synthesize and communicate survey results to local mental health policymakers. The only real constraint is in obtaining a large enough stockpile of laptop computers to carry out the surveys on a production basis in a number of counties at a time. The ideal situation would be for SAMHSA to underwrite the purchase of these computers and to allow them to be used on a rotating basis by counties throughout the country to carry out local needs assessment surveys.

Summary

This chapter reviewed the research done on the prevalence and correlates of SMI since that research was first advanced as a concept in PL 102-321. Preliminary prevalence estimates based either on post hoc secondary analysis of previously collected psychiatric epidemiological surveys or on indirect estimates from screening scales suggest that 2.4 to 3.3 percent of the U.S. adult population meet criteria for SMI in any given month and that 5.4 to 7.2 percent do so at some time during the year. The demographic risk profile for SMI includes being female, young or middle-aged, unmarried, and of low socioeconomic status. SMI is significantly related to SUDs, although most people with SMI do not have a co-occurring SUD. A brief screening scale for SMI has been developed and is now in use in two of the three large government health surveys that are carried out on an annual basis in the United States: the NHIS and the NHSDA. The third large annual government health survey, the BRFSS, does not include this screening scale, but we recommend that the screening scale be included in this survey as well. Sophisticated estimation and calibration procedures to convert these screening scale scores into aggregate prevalence estimates as well as individual-level estimates of the predicted probability of SMI have been developed and will soon be applied to a large calibration survey that is currently under way. These data and methods will make it possible to track trends and subgroup differences in the prevalence of SMI with excellent precision as well as to make reliable estimates for each State. County-level estimates, in comparison, cannot be made with precision using these methods. A procedure that addresses this problem was described for conducting periodic inexpensive county-level needs assessment surveys. The technology exists for carrying out these small-area surveys efficiently, but coordination and access to shared computer hardware and infrastructure are needed to make such surveys feasible.

REFERENCES

Bull, C. N., Krout, J. A., Rathbone-McCuan, E., & Shreffler, M. J. (2001). Access and issues of equity in remote/rural areas. *Journal of Rural Health, 17*, 356–359.

Endicott, J., Spitzer, R. L., Fleiss, J. L., & Cohen, J. (1976). The Global Assessment Scale: A procedure for measuring overall severity of psychiatric disturbance. *Archives of General Psychiatry, 33*(6), 766–771.

First, M., Spitzer, R., Gibbon, M., & Williams, J. (1997). *Structured clinical interview for DSM-IV axis I disorders, research version, non-patient edition (SCID-I/NP)*. New York: Biometrics Research, New York State Psychiatric Institute.

Furukawa, T. A., Andrews, G., Slade, T., & Kessler, R. C. (2003). The performance of the K6 and K10 screening scales for psychological distress in the Australian National Survey of Mental Health and Well-Being. *Psychological Medicine*, 33(2), 357–362.

Goldberg, D. P., Oldehinkel, T., & Ormel, J. (1998) Why GHQ threshold varies from one place to another. *Psychological Medicine, 28*(4), 915–921.

Guyatt, G., & Rennie, D. (2001). *User's guide to the medical literature: A manual for evidence-based clinical practice*. Chicago: AMA Press.

Hartley, D., Britain, C., & Sulzbacher, S. (2002). Behavioral health: Setting the rural health research agenda. *Journal of Rural Health, 18*, 242–255.

House, J. S. (2002). Understanding social factors and inequalities in health: 20th century progress and 21st century prospects. *Journal of Health and Social Behavior, 43*(2), 125–142.

Kessler, R. C., Andrews, G., Colpe, L. J., Hiripi, E., Mroczek, D. K., Normand, S.-L. T., et al. (in press). Short screening scales to monitor population prevalences and trends in nonspecific psychological distress. *Psychological Medicine*.

Kessler, R. C., Andrews, G., Mroczek, D., Ustun, T. B., & Wittchen, H. U. (1998). The World Health Organization composite international Diagnostic Interview Short Form (CIDI-SF). *International Journal of Methods in Psychiatric Research, 7*, 171–185.

Kessler, R. C., Barker, P. R., Colpe, L. J., Epstein, J. F., Gfroerer, J. C., Hiripi, E., et al. (2003). Screening for serious mental illness in the general population. *Archives of General Psychiatry, 60*(2), 184–189.

Kessler, R. C., Berglund, P. A., Walters, E. E., Leaf, P. J., Kouzis, A. C., Bruce, M. L., et al. (1999). Population-based analyses: A methodology for estimating the 12-month prevalence of serious mental illness. In Manderscheid, R. W. & M. J. Henderson (Eds.), *Mental Health, United States, 1998* (pp. 99–109). Washington, D.C.: U.S. Government Printing Office.

Kessler, R. C., Berglund, P. A., Zhao, S., Leaf, P. J., Kouzis, A. C., Bruce, M. L. et al. (1996). The 12-month prevalence and correlates of serious mental illness (SMI). In Sonnenschein, M. A. (Ed.), *Mental Health, United States, 1996* (pp. 59–70). Washington, D.C.: U.S. Dept. of Health and Human Services.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., et al. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry, 51*(1), 8–19.

Kessler, R. C., Nelson, C. B., McGonagle, K. A., Edlund, M. J., Frank, R. G., & Leaf, P. J. (1996). The epidemiology of co-occurring addictive and mental disorders: Implications for prevention and service utilization. *American Journal of Orthopsychiatry, 66*, 17–31.

Kessler, R. C., & Walters, E. E. (2003). The National Comorbidity Survey. In Tsaung, M. T., M. Tohen, & G. E. P. Zahner (Eds.), *Textbook in psychiatric epidemiology, second edition*. New York: John Wiley & Sons.

Peirce, J. C., & Cornell, R. G. (1993). Integrating stratum-specific likelihood ratios with the analysis of ROC curves. *Medical Decision Making, 13*(2), 141–151.

Rehm, J., Ustun, T. B., Saxena, S., Nelson, C. B., Chatterji, S., Ivis, F., et al. (1999). On the development and psychometric testing of the WHO screening instrument to assess disablement in the general population. *International Journal of Methods in Psychiatric Research, 8*, 110–123.

Robins, L. N., & Regier, D. A. (Eds.). (1991). *Psychiatric disorders in America: The Epidemiologic Catchment Area study.* New York: The Free Press.

Robins, L. N., Wing, J., Wittchen, H.-U., Helzer, J. E., Babor, T. F., Burke, J. D., et al. (1988). The Composite International Diagnostic Interview: An epidemiologic instrument suitable for use in conjunction with different diagnostic systems and in different cultures. *Archives of General Psychiatry, 45*, 1069–1077.

Schaible, W. (1996). *Indirect estimators in U.S. Federal Programs*. New York: Springer.

Substance Abuse and Mental Health Services Administration. (1993). Final notice establishing definitions for (1) children with a serious emotional disturbance, and (2) adults with a serious mental illness. *Federal Register, 58*(96), 29,422–29,425.

World Health Organization. (2001). *International classification of functioning, disability, and health*. Geneva, Switzerland.